A Missing Policy: Capacity Building for Sharing Scientific Knowledge

Walter Warnick and David Wojick

***Abstract*—Federal research policy focuses on how much should be invested in specific scientific problems and topics. Little attention is given to investment in the capacity needed to share the results of this research. Given the Internet revolution and the rise of i-Science this policy vacuum needs to be filled.**

Search, knowledge sharing, knowledge diffusion, i-Science, knowledge investment.

I. Introduction

Historically, federal research policy has focused on investment in research, with the primary issue being how much should be invested in specific scientific problems and topics. Very little policy attention is given to investment in the capacity needed to share the results of this research. However, given the Internet revolution and the rise of i-Science this policy vacuum needs to be filled.

The U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) has sponsored original research looking into the potential for the Internet sharing of science to accelerate the global pace of discovery and innovation. The results strongly suggest that investment in the capacity for sharing is just as important as investment in research itself. If so, then we need a research sharing investment policy just as much as we need a research investment policy.

We are not alone in drawing this conclusion. For example, the February 11, 2011 issue of Science was dedicated to presenting the case for a new federal investment policy for data libraries. See http://www.sciencemag.org/site/special/data/ Data access and data libraries are just one of several major issues when it comes to investing in the capacity to share scientific information.

Manuscript received March 9, 2011.

W. Warnick is Director of the Office of Scientific and Technical Information, U.S. Department of Energy, http://www.osti.gov.

D. Wojick is Senior Consultant for Innovation at the Office of Scientific and Technical Information, U.S. Department of Energy, http://www.osti.gov.

II. The OSTI Corollary

Established in 1947, the DOE Office of Scientific and Technical Information (http://www.osti.gov/) fulfills the agency's responsibilities to collect, preserve and disseminate scientific and technical information (STI) emanating from DOE R&D activities. OSTI's mission is to advance science and sustain creativity by making R&D findings available to DOE and other researchers and the public. In the words of Section 982 of the Energy Policy Act of 2005, "The Secretary, through the Office of Scientific and Technical Information, shall maintain with the Department publicly available collections of scientific and technical information resulting from research, development, demonstration, and commercial applications supported by the Department."

OSTI already has made important progress. It has championed an aggressive effort on a series of fronts to make authoritative science information more efficiently available to researchers and the public alike. It has played a leading role in developing and adopting cutting-edge web tools such as relevancy ranking, technology that allows search results to be returned in a ranked order relevant to the search query, and federated search (http://www.osti.gov/fedsearch), the simultaneous search of multiple web databases in real time via a single search query, to enhance the diffusion of scientific knowledge.

OSTI today is host to an outstanding national resource, the Science Accelerator (http://www.scienceaccelerator.gov/), which serves to speed up science discovery by providing access to high quality scientific and technical information. OSTI's DOE collection includes the results of DOE's research and development (R&D) projects and programs, descriptions of R&D projects under way or recently completed, major R&D accomplishments, DOE patents and recent research of interest to DOE.

OSTI also conceived and hosts Science.gov (http://www.science.gov/), a gateway to U.S. government information and research results. Science.gov was launched in December 2002, and is an interagency initiative of 18 U.S. government science organizations with 14 federal agencies that contribute content to serve the information needs of the science-attentive citizen, including science professionals, students and teachers, and the business community. Science.gov now is in its fifth generation and searches more than 40 databases and 200 million pages of science information via a single query.

OSTI also hosts Multilingual WorldWideScience.org (http://worldwidescience.org/), a global science gateway that was launched in June 2007 and today searches 400 million pages of science and technology research and development results offered by more than 70 nations. Users can search non-English databases and have search results translated into one of the nine languages currently supported (Chinese, English, French, German, Japanese, Korean, Portuguese, Russian and Spanish).

OSTI's research work has included a series of studies by Luis Bettencourt from Los Alamos National Laboratory and David Kaiser from MIT. They began by applying a model of population dynamics to the historic growth of several emerging scientific fields. These emerging fields range from nanotechnology to bird flu. The results of these studies suggest that the rate of emergence in science is very sensitive to the contact rate among scientists. This work is presented in detail at http://www.osti.gov/innovation/research/.

Bettencourt and Kaiser went on to look at the evolving network structure of collaboration within these emerging fields. Here they found what appears to be a characteristic phase change in the density of collaboration, which occurs when the field takes off. This too suggests that increased contact and communication play a crucial role in emergence. Their report, entitled "The dynamics of scientific discovery: the spread of ideas and structural transitions in collaboration networks," is available at http://www.osti.gov/innovation/research/diffusion/OSTIBettencourtKaiser.pdf.

OSTI was founded on the principle that science progresses only if knowledge is shared; it grew out of the post-World war II initiative to make the scientific research of the Manhattan Project as freely available to the public as possible. Based on the Bettencourt and Kaiser and other research, we have formulated what we call the OSTI Corollary:

III. Accelerating the Sharing of Scientific Knowledge Accelerates the Advancement of Science

In recent years, the advent of the web has opened up the possibility of sharing knowledge with orders of magnitude more people and making it heretofore unimaginably easier to find and use. The possibility of sharing knowledge faster and better led us to posit the OSTI Corollary. If the sharing of knowledge is accelerated, discovery is accelerated. In mathematical parlance, the Corollary might be considered the time derivative of the concept of scientific progress.

If you ask most science policy experts how science can be accelerated, you are likely to hear responses like hire more scientists, provide more labs, provide better instruments, and bigger and faster computers. But decision makers have an additional way to accelerate science to achieve its benefits; namely, invest more in the capacity to share knowledge.

As Isaac Newton said, "If I have seen further than others, it is by standing on the shoulders of giants." But Newton was not alone on those shoulders. Everyone in science, from his day to ours, draws on the work of others. Science is all about the flow of knowledge: New methods, instruments, techniques, concepts, results, questions, data, etc. The flows are endless, complex and in all directions. It is rightly called a diffusion process. Much of our research has focused on this diffusion process and how to speed it up.

If we looked at all the articles, reports, emails and conversations that pass between all the scientists, we could count billions of knowledge transactions every year. This incredible diffusion of knowledge is the very fabric of science. Given that the diffusion of knowledge is central to science, it becomes a major policy issue to see if we can accelerate it. We note in particular that diffusion takes time. Sometimes it takes a long time. Every diffusion process has a speed. Our thesis is that speeding up diffusion will accelerate the advancement of science.

We are especially interested in something we call global discovery. The millions of researchers around the globe are grouped into thousands of communities. While these communities have geographical aspects, a scientific community may best be defined as a group of researchers working on a single scientific problem. The Web of Science indexes about 8,700 journals, representing many different research communities. That's a lot of science to keep up with. Currently it is difficult for researchers, who primarily track journals within their specific discipline, to hear about discoveries made in distant scientific communities.
In fact, diffusion across distant communities can take years. In contrast, within an individual scientific community, internal communication systems are normally quicker. These include journals, conferences, email groups, and other outlets that ease communication.

Many communities use related methods and concepts: mathematics, instrumentation, and computer applications. Thus there is significant potential for diffusion ACROSS communities, including very distant communities. We see this potential for global discovery as a major opportunity within the OSTI corollary.

Diffusion to distant communities takes a long time because it often proceeds sequentially, typically spreading from the community of origin (A) to a neighbor (B), then to community (C), a neighbor of B, and so on. This happens because neighboring communities are in fairly close contact.

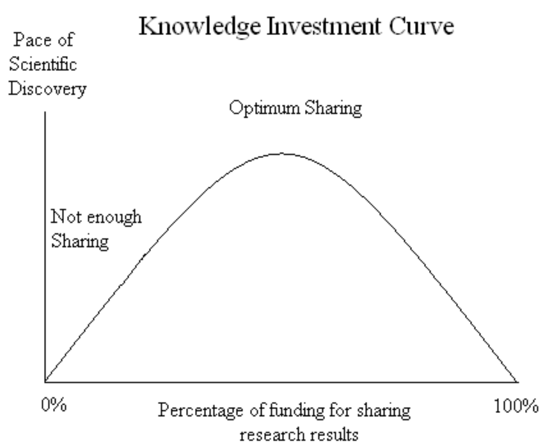
Science will progress faster if this diffusion lag time is diminished. The concept of global discovery is to transform this sequential diffusion process into a parallel process. This means that new knowledge flows directly to distant communities. The goal is to reduce the lag time from years to months and from months to days.

In thinking about how to speed up diffusion across distant communities, we have looked at diffusion research, including computer modeling. We are particularly interested in recent work that applies models of disease dynamics to the spread of scientific ideas. The spread of new ideas in science is mathematically similar to the spread of disease, even though one produces positive results, the other negative. Our goal is to foster epidemics of new knowledge.

IV. The Knowledge Investment Curve

Science advances only if knowledge is shared. Mathematically, this statement implies that the advance of science is a function of both the sharing of research results, as well as doing the original research. In principle, therefore, decision makers face the problem of deciding how much to spend on original research and how much to spend on sharing the knowledge that comes out of research.

Consider the accompanying graph with the x-axis being the fraction of research resources expended on spreading knowledge. The scale ranges from 0% to 100%. The y-axis is the pace of scientific discovery. One can imagine a curve plotting the pace of discovery as a function of the fraction of resources expended on sharing knowledge. A notional curve is shown but we really do not know what the actual curve even looks like. It is a major research question.



When the fraction of resources is 0%, the pace of science advance is zero, as nothing is shared. When the fraction of resources is 100%, the pace of advance is also zero, as nothing is spent on the research itself. In between these endpoints, the plot will have a maximum. The plot is the Knowledge Investment Curve.

While we show a conceptualization of the Knowledge Investment Curve, we know very little about the actual form of this curve, or even how much is currently invested in sharing. Most knowledge sharing activities are not funded directly as budget items. These include writing an estimated one million research papers and reports a year worldwide, as well as finding and reading them. It includes preparing for and participating in conferences, as well as writing and reading emails, blogs, etc. It also includes training postdocs and Ph.D. students, plus an untold number of colleague to colleague conversations.

These myriad activities are centuries old, as old as science itself. What each costs in the aggregate we in many cases have little idea. We do know that scientific journals cost several billion dollars a year, because they depend on a central infrastructure that has a visible price. We also know the budgets of organizations whose purpose is to share knowledge such as the Office of Scientific and Technical Information and sister organizations across the government such as Defense Technical Information Center, The National Library of Medicine, The National Agricultural Library and others.

We also know that the Internet, especially the World Wide Web, is changing the nature of the equation, because the unit cost of sharing is so much less than the traditional means. The Web has made sharing global, or at least potentially so. One of the thrusts of the Office of Scientific and Technical Information is to develop Web tools, like WorldWideScience.org, to promote global discovery. Like the journals, the Web has a visible price in its portals.

We can ask then what the federal investment should be in Web-based science sharing? Conceptually, points on the Knowledge Investment Curve to the left of the optimum imply that the pace of science discovery would be accelerated by increasing the percentage of funding for sharing results. One thing we know is that the investment in sharing is highly uneven across the various sciences, as well as across the various science agencies. The fraction of health science research funding dedicated to sharing knowledge is far greater than for physical and energy sciences. The latter is unlikely to be near the optimum.

Information sharing is an integral part of the R&D process. Thus, decision makers affect the pace of scientific progress when they determine the fraction of R&D dollars dedicated to sharing knowledge. Think of it this way: A program for sharing knowledge derived from scientific research has much in common with a scientific research program itself in that they share the common goal of advancing science. When decision makers of R&D programs discuss optimum funding for research, their decisions are driven by affordability. Similarly, there is an optimum investment in sharing research results as conceptually suggested by the Knowledge Investment Curve. And just as for research itself, the optimum investment is not the minimum.

V. Building Capacity in i-Science

The term “i-Science” refers to Internet-based science, which means science based on radical new access to Information, Ideas, Innovations and Individuals. These are the real i’s in i-

Science.

In the last decade access to new information and facts, especially research results, has increased dramatically. But science is about far more than facts, it is about ideas. New questions, hypotheses and speculations – the engine of scientific progress – circulate faster than ever before. See http://www.osti.gov/ostiblog/home/entry/sharing_results_is_the_engine Moreover, this i-Science revolution is just beginning. A great deal remains to be done to realize its full potential.

Access to innovation is also central to i-Science. Today innovative new methods, approaches and procedures are often quickly available. Most important perhaps is an unparalleled ability to find individuals that can help solve a pressing scientific problem. i-Science is really about people, not just information and ideas.

i-Science is about accelerating scientific progress. But we don't just speed up science when we provide more access to information, ideas and individuals, rather actually we change the way science is done. i-Science works in new ways. Just as the telescope and the microscope transformed the world of science in the 1700s with access to new kinds of information, so access to information and ideas is transforming science today. Life at the scientific frontier is changing, and it will change even more in the years to come. The network structure of science itself is changing. These dramatic changes raise major policy issues, especially for investment policy.

Perhaps the most obvious examples of the way i-Science is changing how science is done is in the highly publicized cases of parallel sharing in big science. Here thousands of scientists share simultaneous access to billion dollar instruments and scientific research data. Astronomy is leading the way here, with things like the Sloan Sky Survey, as is particle physics with the Large Hadron Collider.

As astronomer Ray P. Norris put it, “Fundamental changes are taking place in the way we do astronomy. In twenty years time, it is likely that most astronomers will never go near a cutting-edge telescope, which will be much more efficiently operated in service mode. They will rarely analyze data, since all the leading-edge telescopes will have pipeline processors. And rather than competing to observe a particularly interesting object, astronomers will more commonly group together in large consortia to observe massive chunks of the sky in carefully designed surveys, generating petabytes of data daily. We can imagine that astronomical productivity will be higher than at any previous time. PhD students will mine enormous survey databases using sophisticated tools, cross-correlating different wavelength data over vast areas, and producing front-line astronomy results within months of starting their PhD. The expertise that now goes into planning an observation will instead be devoted to planning a foray into the databases.” (Next-generation Astronomy, 30 September 2010, http://arxiv.org/abs/1009.6027)

Less obvious, but profoundly pervasive, is the spread of i-Science at the bench level. Bear in mind that there are over 8,000 peer-reviewed scientific journals in the world, publishing over a million bench level scientific articles every year. But even this great literature is just the tip of the Internet-based information iceberg. There are hundreds of thousands of Web portals, repositories, publication pages, conference sites, deep web databases, multimedia archives, blogs, and similar sources of vital scientific ideas. Without i-Science no scientist could possibly find useful information in more than a tiny fraction of this great mass of Web accessible thought. But i-Science is making it all findable, and changing the network structure of how science is done in the process.

For example, international collaboration has increased dramatically, because international portals like http://worldwidescience.org/ make it possible for individual scientists to find one another world wide, even if they write in different languages.

Interdisciplinary collaboration has also increased dramatically. Because scientists are no longer confined to just reading journals in their own field, they are finding collaborators across the scientific spectrum. Here one of the most important innovations is the spread of semantic search capabilities that find related articles in far flung sources. The “more-like-this” feature in OSTI’s Information Bridge and the “related articles” feature in Google Scholar are prominent examples. See http://www.osti.gov/ostiblog/home/entry/osti_s_more_like_this

Collaboration is undoubtedly important, but far more common is the act of jumping from idea to idea. Many scientists, in many fields, publish only one or two papers on any given problem, even though they publish dozens or hundreds of papers in their lifetime. This means they are constantly moving from problem to problem, and from idea to idea. Their work is the moving frontier, but at the same time they need to know where everyone around them is. They need to find the new ideas to work on. The diffusion dynamics are incredible. See http://www.osti.gov/ostiblog/home/entry/leaping_concepts_and_global_discovery

This is especially true with university researchers, who in each project are working with graduate and post doctoral students. Every student needs a new topic and in most cases these projects put these students on their lifetime career paths. The better the choice of ideas to pursue the better science will be in the future. I-Science has an essential role to play here, putting students and ideas together.

In short, we are working toward the point where every scientist draws upon the whole world’s scientific knowledge, in a personally efficient manner. This is i-Science. The key concept here may be what we call “findability.” That is, how do we enable every scientist to find what they most need in this sea of published ideas? Making this sea of ideas accessible is just the start. Access must be followed by findability.

Thus we see that faster diffusion means much more than just moving ideas around. It means that scientists do things they would not have otherwise done. They work on different problems, using different methods, with different people, and drawing upon a range of new information formats. The whole structure of scientific activity changes with i-Science.

Science Magazine summed it up nicely in an editorial on "Insights of the Decade:" "In the past 10 years, new ways of gathering, analyzing, storing, and disseminating information have transformed science. Researchers generate more observations, more models, and more automated experimentation than ever before, creating a data-saturated world. The Internet has changed how science is communicated and given nonscientists new opportunities to take part in research. Whole new fields, such as network science, are arising, and science itself is becoming more of a network – more collaborative, more interdisciplinary – as researchers recognize that it takes many minds and varied expertise to tackle complex questions about life, land, and the universe." (www.sciencemag.org, Insights of the Decade, 17 December 2010, page 1612)

As Ray Norris points out, in twenty years things will be far more advanced. But progress does not just happen. In fact i-Science is just getting started, and OSTI is striving to realize its potential. The OSTI Corollary – accelerating the sharing of knowledge accelerates the advancement of science – explains why we at OSTI are constantly striving to share more science with more people faster and more conveniently than before. But there are a number of emerging policy issues with i-Science, most especially with investment policy. Capacity building is not free, even if the Internet is. Sharing knowledge competes with doing research, and sharing knowledge is just as important. The science policy community needs to understand this tradeoff if we are to achieve the promise of i-Science.